

**Amendments to the Specification:**

*Please replace paragraph [0034] with the following rewritten paragraph:*

[0034] Figures 2 and ~~10~~ 11 show an embodiment of the detection step 1 directed to the extraction of text overlays. Note that this embodiment may be combined with a further embodiment discussed below to create a method for detecting both textual and graphic overlays. In Figure 2, the video is scanned in Step 11. Prior to scanning, the video frames may be decomposed into “image hierarchies” according to methods known in the art; this is particularly advantageous in detecting text overlays with different resolutions (font sizes). Scanning here means using a small window (in an exemplary embodiment, 16 x 16 pixels) to scan the image (i.e., each frame) so that all of the pixels in image are processed based on a small window of surrounding pixels. Following scanning 11, the video is subjected to wavelet decomposition 12, followed by feature extraction 13 based on the wavelet decomposition 12. The extracted features are then fed into a neural network processing step 14. In a preferred embodiment, shown in Figure 10, the neural network processing step entails the use of a three-layer back-propagation-type neural network. Based on the features, neural network processing 14 determines whether or not the features are likely to define a textual overlay. This may be followed by further processing 15; for example, in the case in which image hierarchies are used, further processing 15 may entail locating the candidate overlay blocks in the various hierarchy layers and re-integrating the hierarchy layers to restore the original resolution.